

Evaluation and Classification of Leaflet Shape and Size in Wild Soybean

Yiwu Chen and Randall L. Nelson*

ABSTRACT

Two gene pairs are known to affect leaflet shape in soybean [*Glycine max* (L.) Merr.], but only two categories of leaflet shape are routinely used to characterize soybean germplasm. Very little information has been published about leaflet shape and size for wild soybean (*Glycine soja* Sieb. & Zucc.) accessions in the USDA Soybean Germplasm Collection but these characteristics have been shown to be among the most diverse morphological traits in the Chinese *G. soja* collection. The objectives of this research are to evaluate and establish a visual classification system for the variation of leaflet shape and size in the USDA Soybean Germplasm Collection. In 1998 and 1999, 661 wild soybean accessions from the USDA Soybean Germplasm Collection in maturity groups 000 through IX were grown at either Urbana, IL, or Stoneville, MS. Images of fully expanded terminal leaflets were recorded with a digital camera after all accessions were in the reproductive phase. Representative leaflets were selected at approximately two-thirds of the distance from the ground to the top of the plant. Images were analyzed by the computer software Image-Pro Plus (Version 3.0). Six measured or calculated parameters were recorded. The FASTCLUS procedure of SAS was used to define appropriate categories of leaflet shape and size. Length/width ratio and length were chosen to define leaflet shape and leaflet size, respectively. Length/width ratio values ranged from 1.3 to 6.2 and length from 3 to 14 cm. Five categories of leaflet shape are proposed: oval, ovate, lanceolate, linear, and ultra linear; and three classes for leaflet size: small, intermediate, and large. Leaflet shape and leaflet size are associated with geographical origin. Accessions from South Korea were generally smaller than those from China, Japan, and Russia. Accessions from China had more leaflet variation than those from South Korea or Japan but those from Russia were the most variable. Nearly all of the accessions with lanceolate and linear leaflets originated from Russia.

LEAFLET SHAPE is among the most diverse morphological traits of wild soybean (Dong et al., 1999). Since most morphological and pigment traits lack variation in wild soybean, this could be an important trait to help characterize *G. soja* germplasm. Two loci were reported to affect leaflet shape in soybean, *Ln/ln* for ovate leaflet and narrow leaflet (Bernard and Weiss, 1973) and *Lo/lo* for ovate and oval leaflet (Domingo, 1945). Oval leaflet is a rare phenotype and *G. max* germplasm generally is classified either as broad or narrow. Sawada (1988, 1992) proposed the leaflet shape index (ratio of leaflet

length to width) to indicate leaflet shape. He also defined leaflet shape index at 2.6 to differentiate between broad and narrow leaflet shape in Japanese cultivars and found that broad leaflet is a completely dominant trait. Narrow leaflet is associated with an increase in number of seeds per pod in *G. max* (Takahashi, 1934; Domingo, 1945). You et al. (1995) examined 72 lines from China, Japan, and USA for the effects of leaflet shape on seed yield and its components in cultivated soybean and concluded that there were no significant difference between the broad and narrow leaflet lines in average seed yield or number of pods per plant, but significant differences were observed for seed size and number of seeds per pod.

There are several ways to analyze leaflet shape and size data. Frusta et al. (1995) evaluated leaflet shape in 39 cultivated soybeans by principal component analysis based on the elliptic Fourier coefficients. Hill (1998) used a modified Landmark Eigenshape Analysis (LEA) software developed by Dr. J. F. Reid of the Agricultural Engineering Department of the University of Illinois to analyze the images of leaflet shape of *Glycine tomentella* Hayata accessions. The most recently developed Image-Pro Plus software (Media Cybernetics, Silver Spring, MD) has been widely used in various areas of biology (<http://www.mediacy.com>; verified 28 October 2003). Kosina and Wasylikowa (1999) analyzed morphological and anatomical features of plants including hazel nut (*Corylus avellana* L.) shells, primula (*Primula praenitens* Ker Gawl.) and crocus species (Iridaceae) pollen, and *Triticum dicoccon* Schrank. spikelets with the use of Image-Pro Plus software. The greatest advantage of using a computerized image analysis is that many parameters can be measured or calculated with high accuracy and simultaneously analyzed.

Little characterization of leaflet size and shape in *G. soja* has been done. Fukui and Sunaga (1978) investigated morphological variation among 100 accessions of *G. soja* collected from Siberia, northeastern China, South Korea, and Japan and found that leaflet shape was significantly associated with geographical origin. Zheng and Chen (1980) evaluated 478 accessions of *G. soja* from Jilin province of China and proposed four categories for leaflet shape (ovate, long ovate, lanceolate and linear) and three classes for leaflet size (small, intermediate and large) in *G. soja*. Four types of leaflet shape also were employed to evaluate *G. soja* germplasm in China (Li, 1990, 1994). Only 197 of 1104 *G. soja* accessions in the USDA Soybean Germplasm Collection have been evaluated for leaflet length and leaflet shape index (Juvik et al., 1989), and no categories of leaflet shape and leaflet size were defined. The objectives of this research are to evaluate and establish a

Yiwu Chen, Dep. of Crop Sciences, 1101 W. Peabody Dr., University of Illinois, Urbana, IL 61801; Randall, L. Nelson, USDA-Agricultural Research Service, Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, Dep. of Crop Sciences, 1101 W. Peabody Dr., University of Illinois, Urbana, IL 61801. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the USDA or the University of Illinois and does not imply its approval to the exclusion of other products or vendors that may also be suitable. Received 17 Mar. 2003. * Corresponding author (rlnelson@uiuc.edu).

Published in Crop Sci. 44:671–677 (2004).
© Crop Science Society of America
677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: MG, Maturity Group; PI, Plant Introduction; LSD, Fisher's Least Significant Differences.

visual classification system for the variation of leaflet shape and size for wild soybean accessions in the USDA Soybean Germplasm Collection.

MATERIALS AND METHODS

In 1998 and 1999, 498 *G. soja* accessions from USDA Soybean Germplasm Collection in maturity groups (MG) 000 through IV were planted in unreplicated hill plots inside aphid-proof cages at Urbana, IL, but only 279 produced measurable plants in both years. The soybean cultivar Clark and L62-1579, an isoline of Clark with the *ln* allele for narrow leaflet, were planted at Urbana as reference plants. In the same years, 606 accessions in MG V through IX were planted in a hill plots in an open field at Stoneville, MS, but only 382 were successfully grown in both years. There were poor stands in both locations in 1998. Digital images of a fully expanded leaflet at approximately two-thirds of the distance from the ground to the top of the final plant height were recorded each year. Two trifoliate leaves were photographed per plot, but only the terminal leaflet was analyzed.

Images from this experiment were analyzed using the computer software Image-Pro Plus, and six parameters were recorded or calculated: area, length, width, length/width ratio, radius ratio, and roundness. Area includes all pixels within the leaflet perimeter. Length and width is the maximum length and width of the leaflet. The radius ratio is the ratio determined by maximum radius/minimum radius. Radius is the distance between the centroid pixel position of the leaflet and the perimeter. Roundness is calculated as $\text{perimeter}^2 / (4 \cdot \pi \cdot \text{area})$. Circular objects will have a roundness = 1, whereas other shapes will have a roundness >1.0. Length/width ratio is the ratio between the maximum length to width. Length/width ratio is always ≥ 1 .

An analysis of variance (ANOVA) was conducted for each location separately for all parameters. Since not all images could be used for analysis, only 606 accessions had two samplings in 2 yr so GLM was selected for analysis of the unbalanced data set (SAS Institute, 1999). Year and samples were considered random effects, and accessions were considered fixed. Fisher's Least Significant Difference (LSD) at the 1% significance level was used to test the differences among defined leaflet classes. The FASTCLUS procedure of SAS was applied to the data to help determine classes of leaflet types.

RESULTS AND DISCUSSION

Only 606 accessions had useable images for two leaflets in each year and those observations were used for the analysis of variance. There were 661 accessions that were successfully grown in both years and the values collected on those plots were used to compare changes

between years. Over the 2 yr of this research, data were collected on 1021 accessions. All of those observations were used to compare differences among countries and to examine the total range of diversity within the USDA wild soybean collection.

The ranges of the six measured parameters showed that large variation exists among the *G. soja* accessions analyzed. Area, length, and width are all indicators of leaflet size and were all highly correlated (data not shown). Length/width ratio, radius ratio, and roundness, all indicators of leaflet shape, also were highly correlated (data not shown). Length/width ratio was selected to represent leaflet shape because it can be most easily calculated. Both length and width are much easier to measure than total area and either could be selected as an indicator of leaflet size, but narrow leaflets present a problem with either trait. Many of the longest leaflets are also narrower than the mean of the population so classifying leaflet area based on leaflet length will overestimate area for the longest leaflets. A similar problem exists if we use leaflet width since some of the narrowest leaflets are longer than the mean of the population, so leaflet area will likely be underestimated for some of the narrowest leaflets. Since the values for leaflet length are generally larger than those for leaflet width (Table 1) and the range of values for leaflet length (2.6–14.2 cm) is nearly twice as great as leaflet width (1.4–7.7 cm), it will be easier to classify differences visually in leaflet length, so it was selected as an estimate of leaflet area.

The consistency of the parameters from year to year is critical for useful descriptive traits for germplasm evaluation. The ANOVA results showed that the effects of accession, year and the accession \times year interactions are significant for length/width ratio and length in both locations as well as for all MGs with the all P values equal or less than 0.026. The mean values for length/width ratio for 1998 and 1999 were 2.0 and 2.3, and for length were 6.9 and 7.1 cm, respectively. The distribution patterns of length/width ratio means were similar between years with a small shift to higher values in 1999 (Fig. 1) as indicated by the overall means. Among 606 tested accessions, with two leaflet samples in each of the two years, there were 36 entries with no change in length/width ratio between years; 557 entries had an increased value and 13 entries decreased in length/width ratio. Regardless of direction of change, only 29 accessions changed by more than 1.0 unit between years. The 13 accessions with negative changes were evenly

Table 1. Class means of leaflet length/width ratio and length and *t* tests among class means for 661 *G. soja* accessions from the USDA Soybean Germplasm Collection grown in both 1998 and 1999.

Parameter	Class	Cluster centroid by FASTCLUS	Defined range	Class means	No. of accessions
Length/width ratio	Oval	1.6	≤ 2.0	1.8a†	437
	Ovate	2.1	2.1 to 3.0	2.3b	150
	Lanceolate	3.2	3.1 to 4.0	3.5c	43
	Linear	4.2	4.1 to 5.0	4.5d	24
	Ultra linear	5.3	≥ 5.1	5.4e	7
Length (cm)	Small	4.8	≤ 6.0	4.8a	236
	Intermediate	8.1	6.1 to 10.0	7.9b	374
	Large	11.7	≥ 10.1	11.0c	51

† Means with the same letter are not significantly different at the 0.01 probability level. $\text{LSD}_{0.01}$ value for length/width ratio = 0.13; $\text{LSD}(0.01)$ value for length = 0.5.

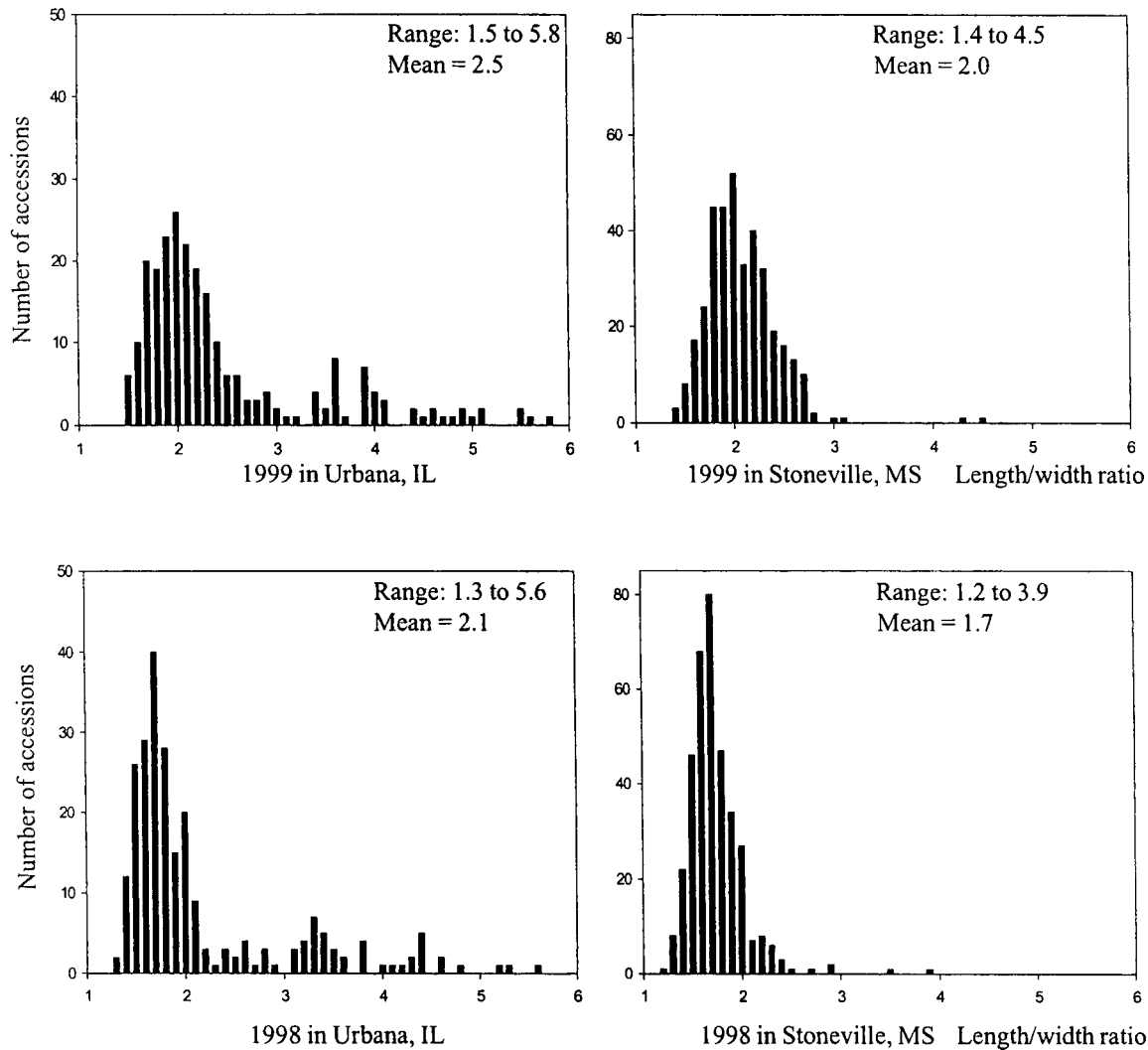


Fig. 1. The distribution of length/width ratio means for *G. soja* accessions grown at Urbana, IL and Stoneville, MS, in 1998 and 1999.

distributed between the two locations and among the MGs. The leaflet length changes were similar to what was observed with the length/width ratio. The means of differences between accessions in the two years was 2.2 cm, and the range of difference was 0 to 10.2 cm in absolute values. There were 16 accessions that did not change between years, 349 accessions that increased and 241 accessions that decreased. As with length/width ratio, the 241 accessions with negative changes were equally distributed across both locations and among all MGs. There are 86 accessions with differences over 4.0 cm among 606 tested accessions. Leaflet length is more sensitive to environmental effects than leaflet shape.

On the basis of our observations, differences of less than 0.5 of length/width ratio value or 3 cm for length would be too small to detect accurately in a visual classification system, but differences approximately 20% of that size were statistically significant in our ANOVA. Both leaflet length/width ratio and length are affected by environmental conditions, but these data indicate that changes among environments are not large enough to negate these parameters as useful in characterizing *G. soja* accessions.

Clark, with the *Ln* gene, is considered to have an ovate leaflet shape in *G. max* and in this research had a leaflet length/width ratio score of 1.8. L62-1579, the isolate of Clark, with the *ln* gene, has the narrow leaflet shape and in the study had a leaflet length/width ratio score of 3.6. The length/width ratio values for the 661 *G. soja* accessions grown and measured in both years ranged from 1.3 to 6.2. The variation observed among the *G. soja* accessions (Fig. 2) is greater than that explained by the known genetics in *G. max*. There were five clusters generated by the FASTCLUS procedure using the length/width ratios, and the cluster centroids were 1.6, 2.1, 3.2, 4.2, and 5.3. Three clusters were formed for leaflet length with centroids of 4.8, 8.1, and 11.7 cm (Table 1). Our results are similar to the results from Zheng and Chen (1980). They surveyed 478 *G. soja* accessions from Jilin province of China and proposed four categories for leaflet shape and three classes for leaflet length in *G. soja*. The length/width ratio means of each of their categories in leaflet shape were 1.8, 2.5, 3.1, and 5.5 and length means were 3.9, 6.8, and 9.6 cm. On the basis of the FASTCLUS results and observations from the field, we defined five categories for leaflet

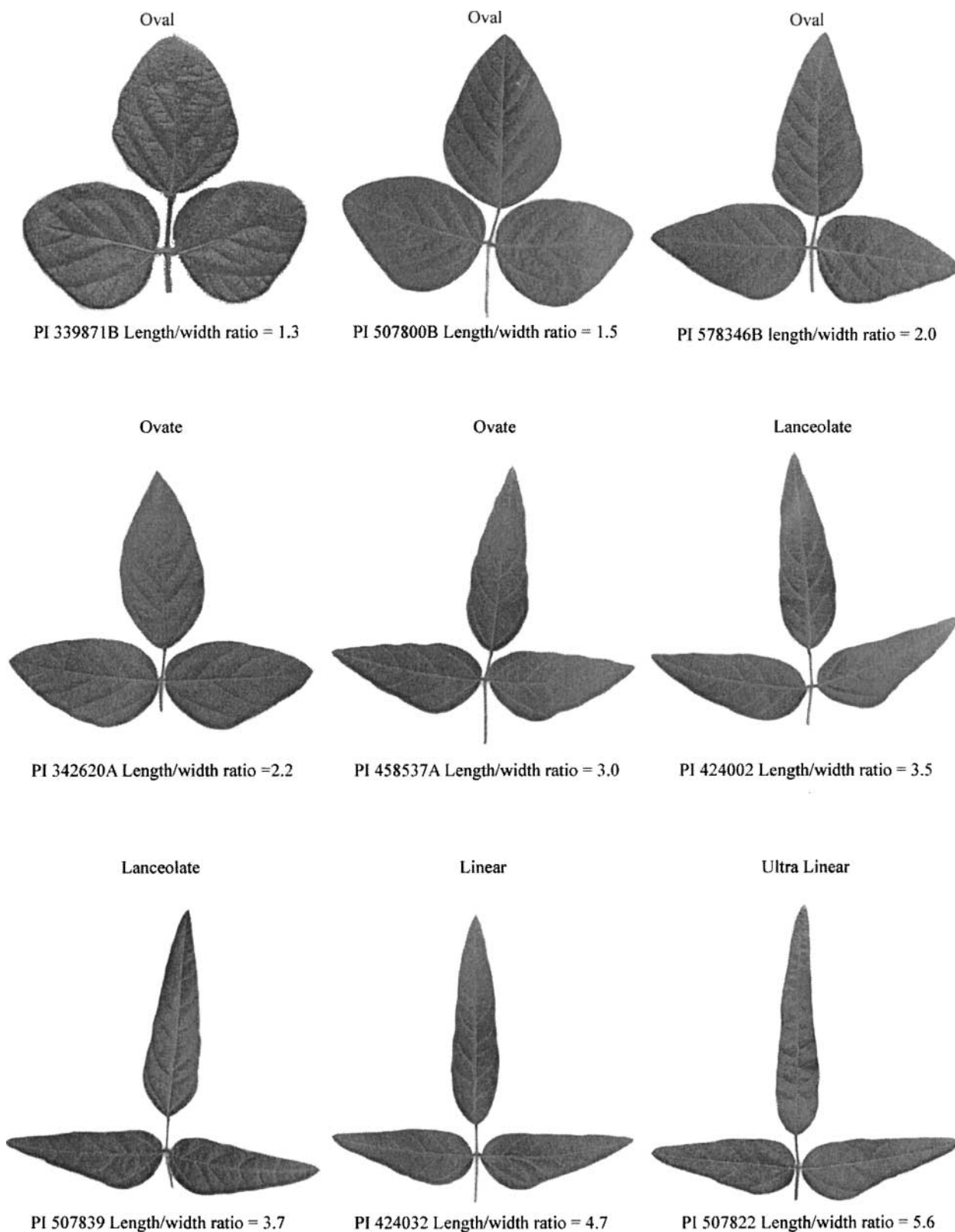


Fig. 2. Range of variation in leaflet shape among *G. soja* accessions in the USDA Soybean Germplasm Collection.

shape by length/width ratio value: less than 2.0 (oval); 2.1 to 3.0 (ovate); 3.1 to 4.0 (lanceolate); 4.1 to 5.0 (linear); and over 5.0 (ultra linear). Three classes are also proposed for leaflet length: small leaflet is less than

6.0 cm in length, intermediate leaflet from 6.1 to 10.0 cm and large leaflet over 10.0 cm. To test for differences among means of each class, *t* tests were used. The results indicated the differences among categories of leaflet

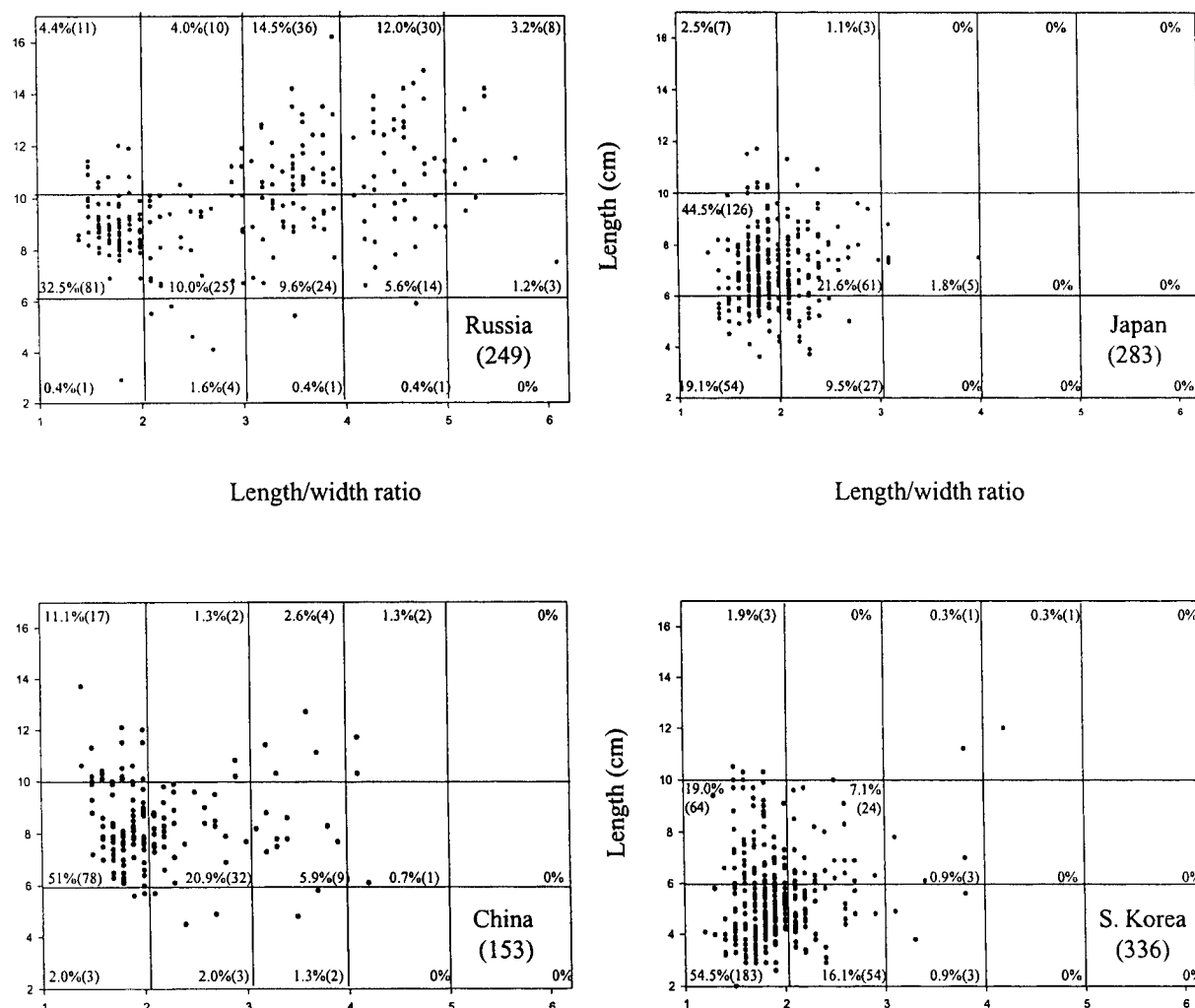


Fig. 3. Scatter plots of *G. soja* accessions in the USDA Soybean Germplasm Collection from four countries separated by five classes of leaflet shape and three classes of leaflet length. Percent and numbers of accessions are given for each category.

shape and leaflet length were highly significant, and means of each class of length/width ratio value and length were very close to the cluster centroids defined by the FASTCLUS procedure (Table 1). The majority of *G. soja* accessions analyzed in this study have oval or ovate leaflet shape and small or intermediate leaflet size.

The means of length/width ratios in MG 000 and 00 were 3.8, while the other MGs were approximately 2.0 except for MG IX, which was 2.5. There are only five accessions in MG IX. There also seems to be an association between leaflet length and maturity. The mean leaflet length of the accessions in MG 000 to IV is over 8.0 cm, whereas in later MGs the mean length is less than 7.0 cm. Since MG IV and earlier were grown in a different location than the MG V and later, it is not totally possible to separate genetic from environmental effects, but the mean length of MG V was much shorter than that of any other group. Thirty-one accessions were classified as linear or ultra linear (length/width ratio >4.1) for leaflet shape. These accessions are all from Russia in MG 000 and 00 except PI 245331, which is from Taiwan and is in MG IX. Of the 43 accessions classified as lanceolate, 35 are from Russia and three from

Northeast China in MG 0 or earlier. There are also two lanceolate accessions from South Korea (MG II and V), two from Japan (MG VII) and one from Taiwan (MG IX). In a survey of morphological variation among 100 accessions of *G. soja* collected from Siberia, northeastern China, South Korea, and Japan, Fukui and Sunaga (1978) found that leaflet shape was significantly associated with geographical origin. In this study, lanceolate leaflet shape was mostly found among entries from Siberia and northeastern China, while South Korean lines were comparatively smaller and oval and Japanese and Chinese lines were comparatively larger and oval or ovate. Our results also indicate a relationship between origin and leaflet shape, but the narrowest leaflets are found at both the northern and southern extremes of the geographical range of *G. soja*.

We also found a relationship between leaflet size and geographical origin. Of the 236 accessions with small leaflet size (<6.0 cm), 163 are from South Korea and 67 from Japan in MG V and VI. Only one is from Russia and the rest are from China. Of the 51 accessions with large leaflet size, 44 are from Russia and five are from Japan. None is from South Korea and only two are

Table 2. Class means for leaflet area and distribution by origin of 1016 *G. soja* accessions from the USDA Soybean Germplasm Collection grown in at least one year. Five accessions from Taiwan are not included.

Class	Cluster centroid by FASTCLUS	Defined range	Class means	No. of accessions	Origin			
					Russia	China	South Korea	Japan
		cm ²						
Small	9.9	<15.0	9.6	421	22	15	255	129
Intermediate	21.3	15.1 to 27.0	20.9	399	118	93	60	128
Large	32.7	>27.1	32.4	196	109	40	21	26

from China. Gerber and Les (1994) compared the leaf morphology among submersed species of *Myriophyllum* (Haloragaceae) from different habitats and geographical distributions. Their results demonstrated that fundamental intraspecific differences in submerged leaf shape were associated with differences in geographic distributions and habitats, and the differences could be explained as adaptations for different nutrient uptake regimes. We do not know what role leaflet size and shape may have in environmental adaptation, but there are definite relationships between the geographical origin and leaflet shape and size in *G. soja* germplasm.

Only 661 accessions were used in the previous analysis because those accessions were successfully grown in both years. The data collected on the remaining 360 accessions were used to classify them into one of the 15 categories on the basis of leaflet size and shape. The distribution of leaflet types by country of origin was visualized with scatter plots of data from 1016 accessions excluding the five accessions from Taiwan (Fig. 3). The predominant class for South Korea was small oval (55%). No other country had such a high percentage of accessions in that class. There were few accessions in the lanceolate or linear classes and even fewer large leaflet types. The distribution of accessions from Japan is similar to that of South Korea, but the predominant class was intermediate oval (45%). There were slightly fewer accessions in the lanceolate or linear classes than from South Korea but more larger leaflet types. The predominant class for the accessions from China also was intermediate oval (52%), but the Chinese accessions were generally larger

and more diverse in shape than those from Japan (Fig. 3). The accessions from Russia seem to have the most diversity, but there are over 100 more accessions from Russia than from China in the USDA Collection. As was noted earlier almost all of the large lanceolate and linear accessions are from Russia. The accessions from Russia are the only group with accessions distributed in 14 of the 15 possible categories. These results provide additional evidence of the association between leaflet shape and size, and geographical origin.

In this classification scheme, we propose to use leaflet length as an estimate of leaflet area knowing that leaflet area could be overestimated for the longest leaflets. To examine the association between leaflet length and leaflet area, we classified leaflet area into three categories, small, intermediate and large by the FASTCLUS procedure (Table 2). The association of leaflet types based on area and origin was similar to that based on length with 75% of the accessions from South Korea in the small class and nearly 45% of the accessions from Russia in the large class (Table 2). The association of leaflet length and leaflet area was investigated by comparing the consistency of the classification between the two methods based on data from all 1021 accessions. Leaflets classified as small were 98% consistent between the two methods, and the large leaflet classes had the least consistency (57%) as was expected (Table 3). The large oval class had complete identity between the two methods, but nearly 60% of the large leaflet accessions, defined by length, in the other categories were only intermediate in terms of actual leaflet area. None of the large leaflet types, defined by length, were in the small leaflet area class. Of those categorized as intermediate on the basis of leaflet length, approximately 60% were in the intermediate area class and nearly equal numbers were underestimated as were overestimated. All of the underestimates were in the oval and ovate categories but the mean leaflet length of these two groups was over 9 cm so they were all very close to the boundary between intermediate and large leaflets. Nearly 75% of those overestimated were in the oval and ovate categories with a mean leaflet length of approximately 6.5, so they were also very close to being in the small leaflet class. These data indicate that using leaflet length will provide a good approximation for leaflet area for *G. soja* germplasm characterization purposes.

There are large and consistent differences for leaflet size and shape that can be used to distinguish among accessions of wild soybean. The system that we have proposed can be used to effectively classify these differences. These results indicate trends in leaflet size and

Table 3. The association between leaflet size estimated by leaflet length and actual leaflet area in 1021 accessions of *G. soja* from the USDA Soybean Germplasm Collection grown in at least one year.

Leaf size estimate (number of accessions)	Leaflet shape	Leaf area		
		Small	Intermediate	Large
Small (336)	Oval	234	7	0
	Ovate	88	0	0
	Lanceolate	6	0	0
	Linear	1	0	0
	Ultra linear	0	0	0
	Sub-total	329 (98%)	7 (2%)	0 (0%)
Intermediate (550)	Oval	29	213	107
	Ovate	45	86	11
	Lanceolate	16	25	0
	Linear	8	7	0
	Ultra linear	1	2	0
	Sub-total	99 (18%)	333 (61%)	118 (21%)
Large (135)	Oval	0	0	38
	Ovate	0	6	5
	Lanceolate	0	23	9
	Linear	0	23	25
	Ultra linear	0	6	0
	Sub-total	0 (0%)	58 (43%)	77 (57%)

shape associated with geographical origin. As additional data are collected on new wild soybean accessions obtained by the USDA Soybean Germplasm Collection and as other collections are classified, the strength of these associations can be further tested and perhaps evolutionary explanations can be found.

ACKNOWLEDGMENTS

The authors thank Dr. Thomas Kilen, USDA-ARS, for assistance in setting up the field experiments in Mississippi and Dr. Lingying Zhao, University of Illinois, for assistance in analyzing the leaflet shape data.

REFERENCES

- Bernard, R.L., and M.G. Weiss. 1973. Qualitative genetics. p. 117–154. In B.E. Caldwell (ed.) *Soybean: Improvement, production, and uses*. Agron. Monogr. 16. ASA, Madison, WI.
- Domingo, W.E. 1945. Inheritance of number of seeds per pod and leaflet shape in the soybean. *J. Agric. Res.* 70:251–268.
- Dong, Y.S., H. Sun, B. Zhuang, L. Zhao, and M. He. 1999. The genetic diversity in annual wild soybean. p. 147–155. In *Proceed. World Soybean Research Conference VI*, Chicago, IL.
- Fukui, J., and S. Sunaga. 1978. Comparative investigation on morphological interstrain variation among Siberian (USSR), northeastern Chinese, South Korean and Japanese strains of wild soybean, *Glycine soja* Sieb. and Zucc. *J. Faculty Agric. Iwate Univ.* 14:81–94.
- Frusta, N., S. Ninomiya, N. Takahashi, H. Ohmori, and Y. Ukai. 1995. Quantitative evaluation of soybean (*G. max* [L.] Merr.) leaflet shape by principal component scores based on elliptic fourier descriptor. *Breed. Sci.* 45:315–320.
- Gerber, D.T., and D.H. Les. 1994. Comparison of leaf morphology among submersed species of Myriophyllum (Haloragaceae) from different habitats and geographical distributions. *Am. J. Bot.* 81: 973–979.
- Hill, J.H. 1998. Morphological and biochemical analysis of variation in diploid *Glycine tomentella* Hayata (2n = 38, 40). Ph.D. diss. (Diss Abstr Inter Vol: 60–03, Section: B, page: 0881). University of Illinois at Urbana-Champaign.
- Juvik, G., R.L. Bernard, R.Z. Chang, and J.F. Cavins. 1989. Evaluation of the USDA Wild Soybean Germplasm Collection: Maturity Groups 000 to IV (PI 65.549 to PI 483.464). USDA Tech Bull No. 1761.
- Kosina, R., and K. Wasylikowa. 1999. Application of computerized image analysis for description of fossil plant remains. p. 317–332. In *Polish Botanical Studies, Guidebook series*. No.23. PAN, Krakow.
- Li, F.S. 1990. Chinese *G. soja* collection catalog. China Agricultural Press, Beijing, China.
- Li, F.S. 1994. Chinese *G. soja* collection catalog (continued). China Agricultural Press, Beijing, China.
- SAS Institute. 1999. SAS/STAT user's guide, Version 8.0, First ed., SAS Inst., Inc., Cary, NC.
- Sawada, S. 1988. Inheritance of leaflet shape in soybean. *Soybean Genet. Newsl.* 15:61–65.
- Sawada, S. 1992. Time of determination and variations within and between plants in leaf shape of soybean. *Jpn. J. Crop Sci.* 61:96–100.
- Takahashi, N. 1934. Linkage relation between the genes for the forms of leaves and the number of seeds per pod of soybean. *Jpn. J. Genet.* 9:208–225.
- You, M.D., Y.B. Liu, T.J. Zhou, and J.Y. Gai. 1995. Effects of leaf shape on seed yield and its components in soybeans. *Soybean Genet. Newsl.* 22:66–70.
- Zheng, H.Y., and H.D. Chen. 1980. A preliminary study on the resources of wild soybean in Jilin province. *Sci. Agric. Sinica.* 13:26–32.